

#19



1600

RAW SEQUENCE LISTING

DATE: 02/12/2003

PATENT APPLICATION: US/09/858,332B

TIME: 10:35:04

Input Set : D:\SEQLIST.txt

Output Set: N:\CRF4\02112003\I858332B.raw

RECEIVED

FEB 19 2003

TECH CENTER 1600/2900

4 <110> APPLICANT: Tchaga, Grigory S.
 5 Jokhadze, George
 7 <120> TITLE OF INVENTION: Metal Ion Affinity Tags and Methods for
 8 Using the Same
 11 <130> FILE REFERENCE: CLON-056CIP
 13 <140> CURRENT APPLICATION NUMBER: US 09/858,332B
 14 <141> CURRENT FILING DATE: 2001-05-15
 16 <150> PRIOR APPLICATION NUMBER: 09/404,017
 17 <151> PRIOR FILING DATE: 1999-09-23
 19 <150> PRIOR APPLICATION NUMBER: 60/101,867
 20 <151> PRIOR FILING DATE: 1998-09-25
 22 <160> NUMBER OF SEQ ID NOS: 23
 24 <170> SOFTWARE: FastSEQ for Windows Version 4.0
 26 <210> SEQ ID NO: 1
 27 <211> LENGTH: 16
 28 <212> TYPE: PRT
 29 <213> ORGANISM: Artificial Sequence
 31 <220> FEATURE:
 32 <223> OTHER INFORMATION: affinity peptide
 34 <400> SEQUENCE: 1
 35 His Leu Ile His Asn Val His Lys Glu Glu His Ala His Ala His Asn
 36 1 5 10 15
 39 <210> SEQ ID NO: 2
 40 <211> LENGTH: 18
 41 <212> TYPE: PRT
 42 <213> ORGANISM: Artificial Sequence
 44 <220> FEATURE:
 45 <223> OTHER INFORMATION: affinity peptide
 47 <400> SEQUENCE: 2
 48 His Asp Asp His Asp Asp His Asp Asp His Asp Asp His
 49 1 5 10 15
 50 Asp Asp
 54 <210> SEQ ID NO: 3
 55 <211> LENGTH: 16
 56 <212> TYPE: PRT
 57 <213> ORGANISM: Artificial Sequence
 59 <220> FEATURE:
 60 <223> OTHER INFORMATION: affinity peptide
 62 <400> SEQUENCE: 3
 63 His Glu Glu His Glu Glu His Glu Glu His Glu Glu His
 64 1 5 10 15
 67 <210> SEQ ID NO: 4
 68 <211> LENGTH: 18

P.6

ENTERED

RAW SEQUENCE LISTING

DATE: 02/12/2003

PATENT APPLICATION: US/09/858,332B

TIME: 10:35:04

Input Set : D:\SEQLIST.txt

Output Set: N:\CRF4\02112003\I858332B.raw

69 <212> TYPE: PRT
70 <213> ORGANISM: Artificial Sequence
72 <220> FEATURE:
73 <223> OTHER INFORMATION: affinity peptide
75 <400> SEQUENCE: 4
76 His Asp Glu His Asp Glu His Asp Glu His Asp Glu His
77 1 5 10 15
78 Asp Glu
82 <210> SEQ ID NO: 5
83 <211> LENGTH: 18
84 <212> TYPE: PRT
85 <213> ORGANISM: Artificial Sequence
87 <220> FEATURE:
88 <223> OTHER INFORMATION: affinity peptide
90 <400> SEQUENCE: 5
91 His Glu Asp His Glu Asp His Glu Asp His Glu Asp His
92 1 5 10 15
93 Glu Asp
97 <210> SEQ ID NO: 6
98 <211> LENGTH: 4
99 <212> TYPE: PRT
100 <213> ORGANISM: Artificial Sequence
102 <220> FEATURE:
103 <223> OTHER INFORMATION: enterokinase cleavage site
105 <400> SEQUENCE: 6
106 Ile Glu Gly Arg
107 1
110 <210> SEQ ID NO: 7
111 <211> LENGTH: 6
112 <212> TYPE: PRT
113 <213> ORGANISM: Artificial Sequence
115 <220> FEATURE:
116 <223> OTHER INFORMATION: a factor Xa cleavage site
118 <400> SEQUENCE: 7
119 Leu Val Pro Arg Gly Ser
120 1 5
123 <210> SEQ ID NO: 8
124 <211> LENGTH: 8
125 <212> TYPE: PRT
126 <213> ORGANISM: Artificial Sequence
128 <220> FEATURE:
129 <223> OTHER INFORMATION: a thrombin cleavage site
131 <400> SEQUENCE: 8
132 His Pro Phe His Leu Val Ile His
133 1 5
136 <210> SEQ ID NO: 9
137 <211> LENGTH: 10
138 <212> TYPE: PRT
139 <213> ORGANISM: Artificial Sequence

RAW SEQUENCE LISTING

DATE: 02/12/2003

PATENT APPLICATION: US/09/858,332B

TIME: 10:35:04

Input Set : D:\SEQLIST.txt

Output Set: N:\CRF4\02112003\I858332B.raw

```

141 <220> FEATURE:
142 <223> OTHER INFORMATION: .a renin cleavage site
144 <400> SEQUENCE: 9
145 Cys Tyr Pro Tyr Asp Val Pro Asp Tyr Ala
146 1      5      10
149 <210> SEQ ID NO: 10
150 <211> LENGTH: 8
151 <212> TYPE: PRT
152 <213> ORGANISM: Artificial Sequence
154 <220> FEATURE:
155 <223> OTHER INFORMATION: an immunological tag
157 <400> SEQUENCE: 10
158 Asp Tyr Lys Asp Asp Asp Asp Lys
159 1      5
162 <210> SEQ ID NO: 11
163 <211> LENGTH: 11
164 <212> TYPE: PRT
165 <213> ORGANISM: Artificial Sequence
167 <220> FEATURE:
168 <223> OTHER INFORMATION: an immunological tag
170 <400> SEQUENCE: 11
171 Cys Glu Gln Lys Leu Ile Ser Glu Glu Asp Leu
172 1      5      10
175 <210> SEQ ID NO: 12
176 <211> LENGTH: 11
177 <212> TYPE: PRT
178 <213> ORGANISM: Artificial Sequence
180 <220> FEATURE:
181 <223> OTHER INFORMATION: an immunological tag
183 <400> SEQUENCE: 12
184 Cys Glu Gln Lys Leu Ile Ser Glu Glu Asp Leu
185 1      5      10
188 <210> SEQ ID NO: 13
189 <211> LENGTH: 3430
190 <212> TYPE: DNA
191 <213> ORGANISM: Artificial Sequence
193 <220> FEATURE:
194 <223> OTHER INFORMATION: DNA sequence of vector containing cDNA of
195 recombinant enterokinase
197 <400> SEQUENCE: 13
198 gacgaaaggg cctcggtgata cgcctatttt tataggttaa tgtcatgata ataatggttt 60
199 cttagacgtc aggtggcact ttctggggaa atgtgcgcgg aaccctatt tggttatttt 120
200 tctaaataca ttcaaataatg tatccgctca tgagacaata accctgataa atgcttcaat 180
201 aatattgaaa aaggaagagt atgagtattc aacatttccg tgtcgccctt attccctttt 240
202 ttgcggcatt ttgccttcct gtttttgctc acccagaaac gctggtgaaa gtaaaagatg 300
203 ctgaagatca gttgggtgca cgagtgggtt acatcgaact ggatctcaac agcggtgaaga 360
204 tccttgagag ttttcgcccc gaagaacggt ttccaatgat gagcactttt aaagttctgc 420
205 tatgtggcgc ggtattatcc cgtattgacg ccgggcaaga gcaactcggg cgccgcatac 480
206 actattctca gaatgacttg gttgagtact caccagtcac agaaaagcat cttacggatg 540

```

RAW SEQUENCE LISTING

DATE: 02/12/2003

PATENT APPLICATION: US/09/858,332B

TIME: 10:35:04

Input Set.: D:\SEQLIST.txt

Output Set: N:\CRF4\02112003\I858332B.raw

```

207. gcatgacagt aagagaatta tgcagtgcct ccataaccat gactgataac actgcggcca 600
208. acttacttct gacaacgatc ggaggaccga aggagctaac cgcttttttg cacaacatgg 660
209. gggatcatgt aactcgcctt gatcggtggg aaccggagct gaatgaagcc ataccaaacg 720
210. acgagcgtga caccacgatg cctgtagcaa tggcaacaac gttgcgcaaa ctattaactg 780
211. gcgaactact tactctagct tcccggcaac aattaataga ctggatggag gcggataaag 840
212. ttgcaggacc acttctgcgc tcggcccttc cggctggctg gtttattgct gataaatctg 900
213. gagccggtga gcgtgggtct cgcggtatca ttgcagcact ggggccagat ggtaagccct 960
214. cccgtatcgt agttatctac acgacggggg gatcaggcaac tatggatgaa cgaaatagac 1020
215. agatcgctga gataggtgcc tcaactgatta agcattggta actgtcagac caagtttact 1080
216. catatatact ttagattgat ttaaaaacttc atttttaatt taaaaggatc taggtgaaga 1140
217. tcctttttga taatctcatg accaaaaatcc cttaacgtga gttttcgttc cactgagcgt 1200
218. cagaccccggt agaaaagatc aaaggatctt cttgagatcc tttttttctg cgcgtaatct 1260
219. gctgcttgca aacaaaaaaa ccaccgctac cagcgggtgt ttgtttgccg gatcaagagc 1320
220. taccaactct ttttccgaag gtaactggct tcagcagagc gcagatacca aatactgtcc 1380
221. ttctagtgtg gccgtagtta ggccaccact tcaagaactc ttagtcaccg cctacatacc 1440
222. tcgctctgct aatcctgtta ccagtggctg ctgccagtgg cgataagtcg tgtcttaccg 1500
223. gggtggactc aagacgatag ttaccggata aggcgcagcg gtcgggctga acgggggggt 1560
224. cgtgcacaca gcccgacttg gagcgaaacga cctacaccga actgagatac ctacagcgtg 1620
225. agctatgaga aagcgccacg cttcccgaag ggagaaaagg ggacaggatc ccggttaagc 1680
226. gbgcagggtc ggaacaggag agcgacagag ggagcttcca gggggaaacg cctggtatct 1740
227. ttatagtcct gtcgggtttc gccacctctg acttgagcgt cgatttttgt gatgctcgtc 1800
228. agggggggcg agcctatgga aaaacgccag caacgcggcc tttttacggt tcctggcctt 1860
229. ttgctggcct tttgtcaca tgttctttcc tgcgttatcc cctgattctg tggataaccg 1920
230. tattaccgct tttgagttag ctgataccgc tcgccgcagc cgaacgaccg agcgacgca 1980
231. gtcagttagc gaggaagcgg aagagcgccc aatacgcaaa ccgcctctcc ccgcgcgttg 2040
232. gccgattcat taatgcagct ggcacgacag gtttcccagc tggaaaagcg gcagttagcg 2100
233. caacgcaatt aatgtgagtt agctcactca ttaggcaccc caggctttac actttatgct 2160
234. tccggctcgt atgttgtgtg gaattgtgag cggataacaa tttcacacag gaaacagcta 2220
235. tgaccatgat tacgccaagc ttgaaggatc atctcatcca caatgtccac aaagaggagc 2280
236. acgctcatgc ccacaacaag atcgatattg tcggaggagc tgactccaga gaaggagcct 2340
237. ggcttgggt cggtgctctg tatttcgacg atcaacaggt ctgcggagct tctctggtga 2400
238. gcagggattg gctggtgtcg gccgccact gcgtgtacgg gagaaatatg gagccgtcta 2460
239. agtggaagc agtgctaggg ctgcatatgg catcaaatct gacttctcct cagatagaaa 2520
240. ctagggtgat tgaccaaatt gtcataaacc cactacaa taaacggaga aagaacaatg 2580
241. acattgccat gatgcatctt gaaatgaaag tgaactacac agattatata cagcctattt 2640
242. gtttaccaga agaaaatcaa gtttttcccc caggaaagaa ttgttctatt gctggctggg 2700
243. gggcacttat atatcaagg tctactgcag acgtactgca agaagctgac gttccccttc 2760
244. tatcaaatga gaaatgtcaa caacagatgc cagaatataa cattacgga aatatggtgt 2820
245. gtgcaggcta tgaagcagga ggggtagatt cttgtcaggg ggattcaggc ggaccactca 2880
246. tgtgccaa gaacaacaga tggctcctgg ctggcgtgac gtcatttggg tatcaatgtg 2940
247. cactgcctaa tcgcccagg gtgtatgccc gggtcacca gttcacagag tggatacaaa 3000
248. gttttctaca tgagctcgta attagctgag aattcactgg ccgtcgtttt acaacgtcgt 3060
249. gagcctggga aaaccctggc gttaccaca ttaatcgct tgcagcacat cccctttctg 3120
250. ccagctggcg taatagcgaa gagggccgca ccgatcgccc ttcccaacag ttgcgcagcc 3180
251. tgaatggcga atggcgctg atgcggtatt ttctccttac gcatctgtgc ggtatttcac 3240
252. accgcatatg gtgcaactct agtacaatct gctctgatgc cgcatagtta agccagcccc 3300
253. gacacccgcc aacacccgct gacgcgccct gacgggcttg tctgctcccg gcatccgctt 3360
254. acagacaagc tgtgaccgct tccgggagct gcatgtgtca gaggttttca ccgtcatcac 3420
255. cgaaacgcgc

```

RAW SEQUENCE LISTING

DATE: 02/12/2003

PATENT APPLICATION: US/09/858,332B

TIME: 10:35:04

Input Set : D:\SEQLIST.txt

Output Set: N:\CRF4\02112003\I858332B.raw

```

257 <210> SEQ ID NO: 14
258 <211> LENGTH: 279
259 <212> TYPE: PRT
260 <213> ORGANISM: Artificial Sequence
262 <220> FEATURE:
263 <223> OTHER INFORMATION: protein sequence of vector containing cDNA of
264 recombinant enterokinase
266 <400> SEQUENCE: 14
267 Met Thr Met Ile Thr Pro Ser Leu Lys Asp His Leu Ile His Asn Val
268 1 5 10 15
269 His Lys Glu Glu His Ala His Ala His Asn Lys Ile Asp Ile Val Gly
270 20 25 30
271 Gly Ser Asp Ser Arg Glu Gly Ala Trp Pro Trp Val Val Ala Leu Tyr
272 35 40 45
273 Phe Asp Asp Gln Gln Val Cys Gly Ala Ser Leu Val Ser Arg Asp Trp
274 50 55 60
275 Leu Val Ser Ala Ala His Cys Val Tyr Gly Arg Asn Met Glu Pro Ser
276 65 70 75 80
277 Lys Trp Lys Ala Val Leu Gly Leu His Met Ala Ser Asn Leu Thr Ser
278 85 90 95
279 Pro Gln Ile Glu Thr Arg Leu Ile Asp Gln Ile Val Ile Asn Pro His
280 100 105 110
281 Tyr Asn Lys Arg Arg Lys Asn Asn Asp Ile Ala Met Met His Leu Glu
282 115 120 125
283 Met Lys Val Asn Tyr Thr Asp Tyr Ile Gln Pro Ile Cys Leu Pro Glu
284 130 135 140
285 Glu Asn Gln Val Phe Pro Pro Gly Arg Ile Cys Ser Ile Ala Gly Trp
286 145 150 155 160
287 Gly Ala Leu Ile Tyr Gln Gly Ser Thr Ala Asp Val Leu Gln Glu Ala
288 165 170 175
289 Asp Val Pro Leu Leu Ser Asn Glu Lys Cys Gln Gln Gln Met Pro Glu
290 180 185 190
291 Tyr Asn Ile Thr Glu Asn Met Val Cys Ala Gly Tyr Glu Ala Gly Gly
292 195 200 205
293 Val Asp Ser Cys Gln Gly Asp Ser Gly Gly Pro Leu Met Cys Gln Glu
294 210 215 220
295 Asn Asn Arg Trp Leu Leu Ala Gly Val Thr Ser Phe Gly Tyr Gln Cys
296 225 230 235 240
297 Ala Leu Pro Asn Arg Pro Gly Val Tyr Ala Arg Val Pro Arg Phe Thr
298 245 250 255
299 Glu Trp Ile Gln Ser Phe Leu His Glu Leu Val Ile Ser Glu Phe Thr
300 260 265 270
301 Gly Arg Arg Phe Thr Thr Ser
302 275
305 <210> SEQ ID NO: 15
306 <211> LENGTH: 12
307 <212> TYPE: PRT
308 <213> ORGANISM: Artificial Sequence
310 <220> FEATURE:

```

RAW SEQUENCE LISTING ERROR SUMMARY
PATENT APPLICATION: US/09/858,332B

DATE: 02/12/2003
TIME: 10:35:05

Input Set : D:\SEQLIST.txt
Output Set: N:\CRF4\02112003\I858332B.raw

Please Note:

Use of n and/or Xaa have been detected in the Sequence Listing. Please review the Sequence Listing to ensure that a corresponding explanation is presented in the <220> to <223> fields of each sequence which presents at least one n or Xaa.

Seq#:21; Xaa Pos. 2,3,5,6,7,9

Seq#:23; Xaa Pos. 2,3,5,6,8,9